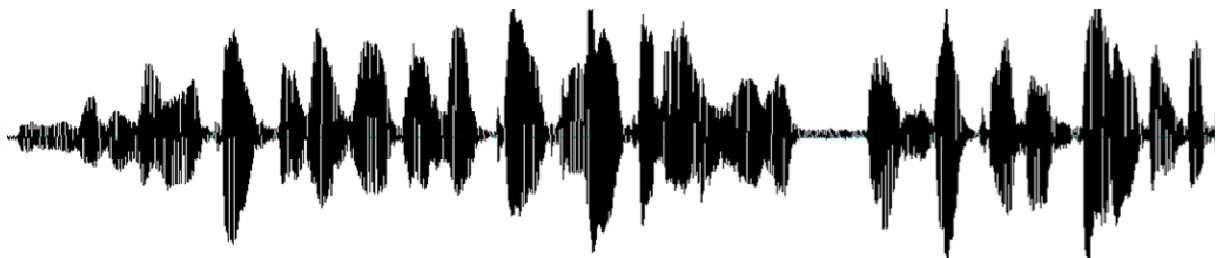


Speech Prosody Satellite Workshop

**Prosodic Prominence:
Perceptual and Automatic Identification**

May 10th, 2010

Doubletree Hotel Magnificent Mile
Ohio - Room
Chicago, Illinois, USA



Organizing Committee

Mathieu Avanzi^{1,2}, Anne Lacheret-Dujour² & Anne Catherine Simon³

¹Neuchâtel, ²Paris Ouest Nanterre, ³Louvain-la-Neuve

mathieu.avanzi@unine.ch ; anne@lacheret.com ; anne-catherine.simon@uclouvain.be

Program

08.00-08.30

Mathieu AVANZI^{1,2}, Anne LACHERET-DUJOUR² & Anne Catherine SIMON³

¹Neuchâtel, Switzerland; ²Paris Ouest Nanterre, France; ³Louvain-la-Neuve, Belgium

Opening session – Prosodic Prominence: An Overview

Session 1: Perceptual and Acoustic Correlates of Prosodic Prominence

08.30 – 9.00

Donna ERICKSON

Showa Music University, Kawasaki City, Japan

An Articulatory Account of Rhythm, Prominence and Phrasal Articulation

9.00-9.30

Denis ARNOLD¹, Petra WAGNER² & Bernd MÖBIUS¹

¹University of Bonn; ²University of Bielefeld, Germany

The Effect of Priming on the Correlations between Prominence Ratings and Acoustic Features

9.30-10.00

Gero KUNTER

Universität Siegen, Germany

Perception of Prominence Patterns in English Nominal Compounds

10.00-10.15

BREAK

Session 2: Automatic Processing

10.15-10.45

Mathieu AVANZI^{1,2}, Anne LACHERET-DUJOUR² & Bernard VICTORRI³

¹Neuchâtel, Switzerland; ²Paris Ouest Nanterre, France; ³Lattice, ENS, Paris, France

A Corpus-based Learning Method for Prominence Detection in Spontaneous Speech

10.45-11.15

Samer AL MOUBAYED, Gopal ANANTHAKRISHNAN & Laura ENFLO

Center for Speech Technology, KTH, Stockholm, Sweden

Automatic Prominence Classification in Swedish

11.15-11.45

Philippe MARTIN

Paris Diderot, France

Prominence Detection without Syllabic Segmentation

11.45-12.15

Giovanni ABETE, Francesco CUTUGNO, Bogdan LUDUSAN & Antonio ORIGLIA

University, Naples, Italy

Pitch Behavior Detection for Automatic Prominence Recognition

12.15-13.45

LUNCH

Session 3: Non-native Speakers and Language Acquisition

13.45-14.15

Andrew ROSENBERG¹ & Julia HIRSCHBERG²

¹Queens College and ²Columbia University, New York, USA

Production of English Prominence by Native Mandarin Chinese Speakers

14.15-14.45

Britta LINTFERT¹ & Bernd MÖBIUS²

¹University of Stuttgart; ²University of Bonn, Germany

Acquisition and Development of Syllabic Prominence in German Speaking Style

14.45-15.00

BREAK

Session 4: Inter-transcriber Consistency in Spontaneous Speech Corpora

15.00-15.30

Tae-Jin YOON

McMaster University, Canada

Speaker Consistency in the Realization of Prosodic Prominence in the Boston University Radio Speech Corpus

15.30-16.00

Mathieu AVANZI^{1,2}, Anne Catherine SIMON³, Jean-Philippe GOLDMAN^{3,4} & Antoine AUCHLIN⁴

¹Neuchâtel, Switzerland; ²Paris Ouest Nanterre, France; ³Louvain-la-Neuve, Belgium, ⁴Genève, Switzerland

C-PROM. An annotated Corpus for French Prominence Studies

16.00-16.30

Massimo MONEGLIA¹, Tommaso RASO², Maryualê MALVESSI-MITTMANN², Heliana MELLO²

¹University of Florence, Italy; ²UFMG, Brazil

Challenging the Perceptual Relevance of Prosodic Breaks in Multilingual Spontaneous Speech Corpora: C-ORAL-BRASIL / C-ORAL-ROM

16.30-17.00

Mathieu AVANZI^{1,2}, Anne LACHERET-DUJOUR² & Anne Catherine SIMON³

¹Neuchâtel, Switzerland; ²Paris Ouest Nanterre, France; ³Louvain-la-Neuve, Belgium

Closing Session and Discussion

Abstracts
(Alphabetic order)

Giovanni ABETE, Francesco CUTUGNO, Bogdan LUDUSAN & Antonio ORIGLIA

University, Naples, Italy

Pitch Behavior Detection for Automatic Prominence Recognition

In this paper a non-supervised approach for automatic syllable prominence recognition is presented. Previous research indicates syllable nuclei energy and duration as the main cues to detect prominence. Fundamental frequency has also been investigated in the past but considered secondary or irrelevant for this task. The proposed system uses the energy and the duration of the nucleus while taking into account also the pitch behavior. The algorithm was tested by comparing its results with the annotations of two human experts and we found that this approach has a 5.6% accuracy increase with respect to the system not using the pitch behavior.

Samer AL MOUBAYED, Gopal ANANTHAKRISHNAN & Laura ENFLO

Center for Speech Technology, KTH, Stockholm, Sweden

Automatic Prominence Classification in Swedish

This study aims at automatically classifying levels of acoustic prominence on a dataset of 200 Swedish sentences of read speech by one male native speaker. Each word in the sentences was categorized by four speech experts into one of three groups depending on the level of prominence perceived. Six acoustic features at a syllable level and seven features at a word level were used. Two machine learning algorithms, namely Support Vector Machines (SVM) and memory based Learning (MBL) were trained to classify the sentences into their respective classes. The MBL gave an average word level accuracy of 69.08% and the SVM gave an average accuracy of 65.17 % on the test set. These values were comparable with the average accuracy of the human annotators with respect to the average annotations. In this study, word duration was found to be the most important feature required for classifying prominence in Swedish read speech.

Denis ARNOLD¹, Petra WAGNER² & Bernd MÖBIUS¹

¹University of Bonn; ²University of Bielefeld, Germany

The Effect of Priming on the Correlations between Prominence Ratings and Acoustic Features

In previous research we showed that the priming paradigm can be used to significantly alter the prominence ratings of subjects. In that study we only looked at the changes in the subjects' ratings. In the present study, we analyzed the acoustic parameters of the stimuli used in the priming study and investigated the correlation between prominence ratings and acoustic parameters. The results show that priming has a significant effect on these correlations. The contribution of acoustic features on perceived prominence was found to depend on the prominence pattern. If a dominantly prominent syllable is present in a given utterance, f_0 and intensity contribute most to the perceived prominence, while duration contributes most when no syllable is dominantly.

Mathieu AVANZI^{1,2}, Anne LACHERET-DUJOUR² & Bernard VICTORRI³

¹Neuchâtel, Switzerland; ²Paris Ouest Nanterre, France; ³Lattice, ENS, Paris, France

A Corpus-based Learning Method for Prominence Detection in Spontaneous Speech

The aim of this paper is to present a software tool called ANALOR, which allows semi-automatic prominences detection in spontaneous French. On the basis of a manual annotation made by two experts on a 70-minute long corpus including different regional varieties of French (Belgian, Swiss and metropolitan French) and various discourse-genres (from reading speech to spontaneous conversations), our system conducts a learning-method in order to get the best thresholds for prominence prediction. This procedure allows honing the detection, the constituency between the automatic identification and the human labeling passing from 75.3 without training to 79.1 of f-measure after corpus-based learning.

Mathieu AVANZI^{1,2}, Anne Catherine SIMON³, JEAN-PHILIPPE GOLDMAN^{3,4} & ANTOINE AUCLIN⁴

¹Neuchâtel, Switzerland; ²Paris Ouest, France; ³Louvain-la-Neuve, Belgium, ⁴Genève, Switzerland

C-PROM. An annotated Corpus for French Prominence Studies.

This paper presents C-PROM, an annotated corpus for French prominence studies. The corpus, including different regional varieties of French (Belgian, Swiss and metropolitan French and various discourse-genres (from oral reading to spontaneous conversations) for a total duration of 70 minutes, was annotated by two phonetics experts. The two experts in charge of the coding followed a strict protocol, which takes into account both the previous mistakes encountered by prior research into prominence detection in French and elements of the methodology followed by scholars working on other languages. We conclude by discussing the average consistency between the two transcribers. The results obtained are quite encouraging, as the F-measure between the two annotators reaches 82.8%, and the kappa-score 0.86.

Donna ERICKSON

Showa Music University, Kawasaki City, Japan

An Articulatory Account of Rhythm, Prominence and Phrasal Articulation

This paper examines some articulatory and acoustic characteristics of American English. The results suggest that the jaw may be the articulatory organizer of phrasal rhythm, manifested acoustically through the F2-F1 pattern. Utterance prominence, such as contrastive emphasis, is additionally manifested by increased F0 along with increased duration on the prominent word. The rhythmical organization of the utterance, based on strong-weak jaw opening patterns, may be different from the intonational organization involving pitch accents/ boundary strengths. American English prosody might be best described using a parallel system involving both a rhythm system based on articulation, and an intonational system involving pitch notations.

Gero KUNTER

Universität Siegen, Germany

Perception of Prominence Patterns in English Nominal Compounds

This paper investigates prominence patterns in English noun-noun compounds. A perception experiment is presented in which naive listeners rated the prominence relation between the two elements. It is found that either the first or the second element of the stimuli is perceived as more prominent, and that the distinction can be considered categorical. Pertinent statistics are discussed that can assess the reliability of different raters. It is shown that right prominence is more difficult to perceive by many listeners than left prominence.

Britta LINTFERT¹ & Bernd MÖBIUS²

¹University of Stuttgart; ² University of Bonn, Germany

Acquisition and Development of Syllabic Prominence in German Speaking Style

An investigation of the acoustic correlates of word stress in infant polysyllabic vocalization was carried out on the basis of data from 6 German-learning infants between 7 and 36 months of age in order to describe the development of word stress in German. The data were analyzed with respect to duration, intensity, fundamental frequency (F0), as well as vowel quality parameters describing the time and degree of opening of the glottis, the slope of the spectrum and glottal leakage. Even with beginning of babbling children are able to produce different stress patterns. However, the implementation and usage of the parameters that contribute to marking word stress appear to be inconsistent. Each measurable acoustic parameter for marking word stress can be observed, and with advancing age the use of these parameters is very variable and depends on the individual child. The most important cue to mark different stress patterns is to learn to reduce the acoustic parameters for the production of unstressed.

Massimo MONEGLIA¹, Tommaso RASO², Maryualê MALVESSI-MITTMANN², Heliana MELLO²

¹University of Florence, Italy; ²UFMG, Brazil

Challenging the Perceptual Relevance of Prosodic Breaks in Multilingual Spontaneous Speech Corpora: C-ORAL-BRASIL / C-ORAL-ROM

A Corpus of Brazilian Portuguese (BP) will join C-ORAL-ROM [1] adopting the same corpus design and prosodic annotation schema. The inter-rater agreement concerning the annotation of terminal and non terminal breaks by both experts and non experts is studied and compared with the early C-ORAL-ROM results [2]. Although the overall prominence of prosodic breaks is confirmed ($K > 0.80$) the inter-rater agreement for terminals turns out satisfactory only for the experts (0.76). Moreover the annotation of non terminal breaks shows low reliability and suffers of language specific factors connected to the rhythmic structure of BP [3]. The paper focuses on the qualitative analysis of the language contexts types determining the low inter-rater agreement and highlights both language specific and general factors which interact with perceptual prominence of prosodic breaks in BP.

Philippe MARTIN

Paris Diderot, France

Prominence Detection without Syllabic Segmentation

Detection of prominence, whether automatically or manually through perception tests, is pivotal in the interpretation of data in a prosodic theoretical framework. This is particularly true for French, where phonologically stressable syllables are not necessarily stressed. To assert a prominence character to syllables is mandatory to evaluate prosodic theories, especially those which predict the phonetic features of melodic contours (rise, fall, height, etc.) located on those syllables. Some algorithms are already available to detect prominent syllables automatically, but most involve a precise segmentation of speech into syllables, vowels and consonants, a task which generally requires a reasonable good quality of recording, exempt from background noise and echo. In order to avoid the problematic segmentation into phonetic units, we propose here an algorithm for prominence detection operating differently and based on readily available phonetic properties of speech, at the exemption of spectral properties.

Andrew ROSENBERG¹ & Julia HIRSCHBERG²

¹Queens College and ²Columbia University, New York, USA

Production of English Prominence by Native Mandarin Chinese Speakers

Native-like production of intonational prominence is important for spoken language competency. Non-native speakers may have trouble producing prosodic variation in a second language (L2) and thus, problems in being understood. By identifying common sources of production error, we will be able to aid in the instruction of L2 speakers. In this paper we present results of a production study designed to test the ability of Mandarin L1 speakers to produce prominence in English. Our results show that there are some consistent differences between the L1 and L2 speakers in the use of pitch to indicate prominence, as well as in the accenting of phrase-initial tokens. We also find that we can automatically detect prominence on Mandarin L1 English with 87.23% and an f-measure of 0.866 if we train a classifier with annotated Mandarin L1 English data. Models trained on native English speech can detect prominence in Mandarin L1 English with an accuracy of 74.77% and f-measure of 0.824.

Tae-Jin YOON

McMaster University, Canada

Speaker Consistency in the Realization of Prosodic Prominence in the Boston University Radio Speech Corpus

An analysis is presented on the rate of inter-speaker consistency in the way multiple speakers realize prosodic events when they read the same scripts. The analysis is made on the Boston University Radio Speech Corpus (BURSC). The BURSC consists of data from five speakers (3 female and 2 male), each reading the same scripts that comprise more than 110 different sentences. The design of the corpus, thus, proves to be a useful basis on which we can measure the degree of speaker variation or speaker consistency in prosodic realization. A pair-wise comparison of inter-speaker consistency is made regarding the rendition of prosodic prominence. The results indicate that the average rate of consistency on the presence or absence of pitch accent is 89.81%. An average consistency of 72.17% is achieved for the rate of consistency for the types of the pitch accent. The finding implies that there is a constraint that is imposed on an utterance by speakers regarding prosodic prominence placement, as well as certain degree of variation between speakers in rendering prosodic prominence.